(2)

AD-A236 893

Classification of Complex Sounds

Semiannual report to the
Office of Naval Research
Grant #ONR N00014-91-J-1122

Bruce G. Berg
Psychoacoustics Laboratory
Department of Psychology
University of Florida
Gainesville, FL 32611
(904) 392-1608

1991

91-01777

91 6 11 056

Most of the work completed during the first semiannual period (since November 1, 1991) involves the use of the COSS analysis to estimate spectral weights in several profile analysis tasks. These tasks include: (1) discrimination of broadband "rippled spectra", (2) discrimination of narrowband spectra, and (3) estimation of "temporal-spectral weights" using a three-tone profile with different durations. This report also includes a description of anticipated work during the next semiannual period.

A. Discrimination of broadband stimuli: rippled profiles

In most profile tasks, listeners discriminate a sound having a flat spectrum from one with a "spectral bump", that is, an increment in the intensity of a single spectral component. Berg and Green (1990) show that the estimated spectral weights for listeners are close to the optimal weights derived from the channels model of Durlach, Braida, and Ito (1986). In other words, listeners are reasonably efficient at integrating information from different spectral regions (i.e. critical bands) in this task. Here, we consider whether or not this finding can be generalized to the detection of spectral changes which are more complex. Rather than increasing the intensity of a single component, systematic changes in relative intensity are imposed on all spectral components (ranging from 200 Hz to 5000 Hz). The shapes of the amplitude spectra for these complex sounds have a sinusoidal appearance, and are called "rippled spectra". (See Fig. 1 in the enclosed manuscript, Berg and Green, (1991); Discrimination of complex spectra: Spectral weights and performance efficiency.) In the case of a one-cycle, sinusoid ripple, for instance, the amplitudes of low frequency components are increased, whereas the amplitudes of high frequency components are decreased. These spectral changes can be described as "global" changes. In comparison, a four-cycle, sinusoid ripple produces a spectrum for which adjacent components (along the frequency axis) are alternatively increased and decreased in level. Spectral changes in this case can be described as "local" changes. One finding from the current work is that listeners show greater efficiency at integrating information when the spectral changes are global than when the spectral changes are local. Other details about this work are discussed in the enclosed manuscript.

One facet of the experiments concerns the effects of extended training. A condition is used in which a standard, consisting of an eight-component, flat spectrum is discriminated from a "signal" for which the amplitude of the second and sixth components are decreased by Δa, and the amplitude of the fourth and eight components are increased by Δa (-sin2 condition). Spectral weights obtained during a typical period of training (about 5000 trials) are shown for three listeners in the panels on the right side of Fig. 1 (attached). The panels on the left side show the spectral weights following an additional 10,000 trials. For each set of weights, we calculate a performance measure, η_{wgt}, which quantifies the efficiency of the observed weights relative to ideal weights. For two listeners, K and G, η_{wgt} increases with extended training, whereas for the third listener, η_{wgt} decreases following the additional training. Note that the final weight estimates for G are remarkably close to ideal weights.

B. Discrimination of narrow band spectra.

Our research has shown that the auditory cues which listeners use to discriminate two spectral profiles are dependent on spectral bandwidth. For wide band profiles, evidence suggests that listeners make across channel, level comparisons (as described in Berg and Green, 1990). For narrower bandwidths, spanning about two to three critical bands, evidence suggests that listeners use differences in spectral pitch in order to discriminate the two sounds. For spectral profiles with bandwidths less than a critical band, we believe that listeners base their decisions on timbre or "roughness". Some of this work, particularly for pitch cues, is discussed in the enclosed paper by Berg, et al. (first draft; will be submitted to J. Acoust. Soc. Am.).

The significance of this work is that it demonstrates the remarkable adaptability of listeners in discriminating complex sounds. The channels model, which is appropriate for wide band spectra, predicts that performance should be quite poor when the spectral bandwidth is less than a critical band. Contrary to this expectation, we found that listeners detect spectral changes in a stimulus with a 20 Hz bandwidth about as well as they discriminate spectral changes in a broadband stimulus. Obviously, listeners must base their decisions on some other auditory cue (e.g. roughness) for narrow bandwidths.

Much of this work has been guided by our finding that the pattern of spectral weights is a function of stimulus bandwidth or stimulus configuration. With three-tone profiles, for instance, there are profound differences among spectral weight estimates for a wideband profile, a 160 Hz wide profile, and a 20 Hz wide profile (all centered about 1000 Hz). Evidence suggests that these spectral weight differences reflect differences in the auditory cues which are used by listeners. In other words, we believe that differences in the weighting functions reflect differences in underlying computational mechanisms. In order to account for these data, we have developed two models for discriminating narrow-band spectra. The first is a modification of Feth's (1974) EWAIF model (discussed extensively in the enclosed manuscript.) The second model entails a calculation based on the spectrum of the temporal envelope. For example, consider a stimulus consisting of the three components at 980 Hz, 1000 Hz and 1020 Hz. For the standard, all three components have the same amplitude, and for the signal, the amplitude of the 1000-Hz component is increased. The spectrum of the envelope (ignoring the DC component) consists of two components at 10 Hz and 20 Hz. The ratio of the Fourier coefficients for these two components is used as a decision statistic to discriminate the two sounds. Performance levels attained by using this computation have thus far greatly exceeded other envelope-based computations. An important property of this computation is that it is unaffected by changes in the overall level of the sound (Recall that a 20-dB variation in level is used in these tasks, which severely limits the usefulness of absolute intensity as a cue). Recently, Dave Green has generalized this model to profiles with more than three components and to tone-in-noise detection tasks. We feel that the model has considerable potential, offering an alternative to traditional energy detection models (which cannot account for the current data).

The mainstay of this work continues to be the spectral analysis. As we have seen, the COSS technique is leading directly to the formulation of new models of underlying computational mechanisms. Moreover, it should be emphasized that spectral weights are behavioral data, and thus may serve as a useful criterion for ruling out competing models of auditory processes. It is noteworthy that the three models which we have examined, the channels model, the EWAIF model, and the model based on the spectrum of the envelope, together provide an account for the different pattern of spectral weights found in the experiments.

C. Spectral-Temporal weights.

The work discussed in this section was done in collaboration with Huanping Dai, an Assistant Research Scientist also working in Dave Green's Psychoacoustic Lab. Much of this work is discussed in the enclosed manuscript, Spectral-Temporal Weights in Profile Detection by Dai and Berg (first draft; will be submitted to the J. Acoust. Soc. Am.). A brief discussion of this work is included here.

A typical stimulus duration in profile analysis tasks is about 500 ms. Since traditional estimates of temporal integration are around 100 ms, one might consider whether the information acquired from different temporal segments of the stimulus contribute equally to a listener's decisions. We examined this issue in several experiments by estimating weights in both the temporal and spectral domains.

In essence, the COSS technique is done by adding a small perturbation to the level of each spectral component and examining the effects of these perturbations on a listener's responses. If a component is relatively important with respect to a listener's responses, then level perturbation in the level of that component should have a large effect on responses. On the other hand, if a component is relatively unimportant, than level perturbations should have little effect on responses. In order to obtain spectral weights, an independent perturbation is added to each component. We extend this technique to obtain spectral-temporal weights. A three-tone profile (200 Hz, 1000 Hz, and 5000 Hz) is 100% amplitude modulated with a period equal to 1/3 the total stimulus duration, thus segmenting the stimulus into three temporal segments. Rather than adding an independent level perturbation to each component, three independent perturbations are added to each component, one during each of the three temporal segments. COSS analysis of the data then yields three weights for each spectral component, one for each of the three temporal segments.

Here, we summarize briefly our findings: (1) If the signal is added to all three temporal segments of the 1000 Hz tone, then all three temporal segments of a component should ideally have the same weight -- unity for the signal component and -0.5 for the two nonsignal components. Generally, listeners do not weight the segments equally, usually giving greater weight to either the initial or final segment. (2) Currently, COSS theory assumes that decisions are based on a weighted sum of the "observations, that is, information across channels is combined linearly. We show that the spectral weight of a component (over the entire stimulus duration) can be recovered by summing the three spectral-temporal weights for that component. Thus, the assumption of linearity is supported empirically. (3) In one experiment, the signal is added to only one temporal segment of the 1000 Hz component. This was done in order to investigate listeners' abilities to adjust their weights. For a stimulus duration of 300 ms, observers appear to be able to adjust their weights in accordance with the signal position. In contrast, for a stimulus duration of 15 ms, two of three observers appear to be unable to adjust their weights in accordance with the signal position. These results are consistent with a temporal integration time of 100 ms, an estimate often reported in the literature. The third observer, however, appears to be able to adjust his weights when the signal position is changed within the 15-ms stimulus, an intriguing result which suggests an integration time of less than 5 ms for this listener.

D. Projected Work

I have recently accepted a position as Assistant Professor in the Cognitive Science Department at the University of California, Irvine, I plan to relocate sometime in July. Most of the summer will be devoted to building a laboratory. I have received enough "startup" funds which, in combination with the funds budgeted in my current grant from ONR, should be sufficient to properly equip the lab. Even though I am leaving his lab, Dave Green will continue to serve as a consultant. In June, I will be attending the 9th International Symposium on Hearing in Carcans, France.

I plan to begin data collection by the end of September. Fortunately, this break in data collection occurs at a convenient time, since I plan to start a different phase of this research project. In a recent paper, Au and Whitlow (1989) report that human listeners are able to make fairly accurate discriminations of dolphin echolocation calls (following a linear transformation of the calls to a frequency range audible to humans); Dr. Au has generously given me a set of these digitized stimuli. My intentions are to use the COSS analysis in an attempt to determine which aspects (temporal and spectral) are used by listeners in order to discriminate these "real-world sounds". The work described in Sec. A suggests that the discriminations may be based on a limited frequency range, whereas the work described in Sec. C offers a method for investigating the utilization of information in the temporal domain. This work will also determine whther the COSS analysis is useful for investigating listeners' discriminations of more complex stimuli.

Theoretical work will continue, including additional investigation of the computational models discussed above. I am also in the process of analyzing additional data (already collected) related to the discrimination of narrow band spectra.

REFERENCES

Au, W.W.L, and Martin, W.M. (1989). Insights into dolphin sonar discrimination capabilities from human listening experiments. J. Acoust. Soc. Am., 86, 1662-1670.

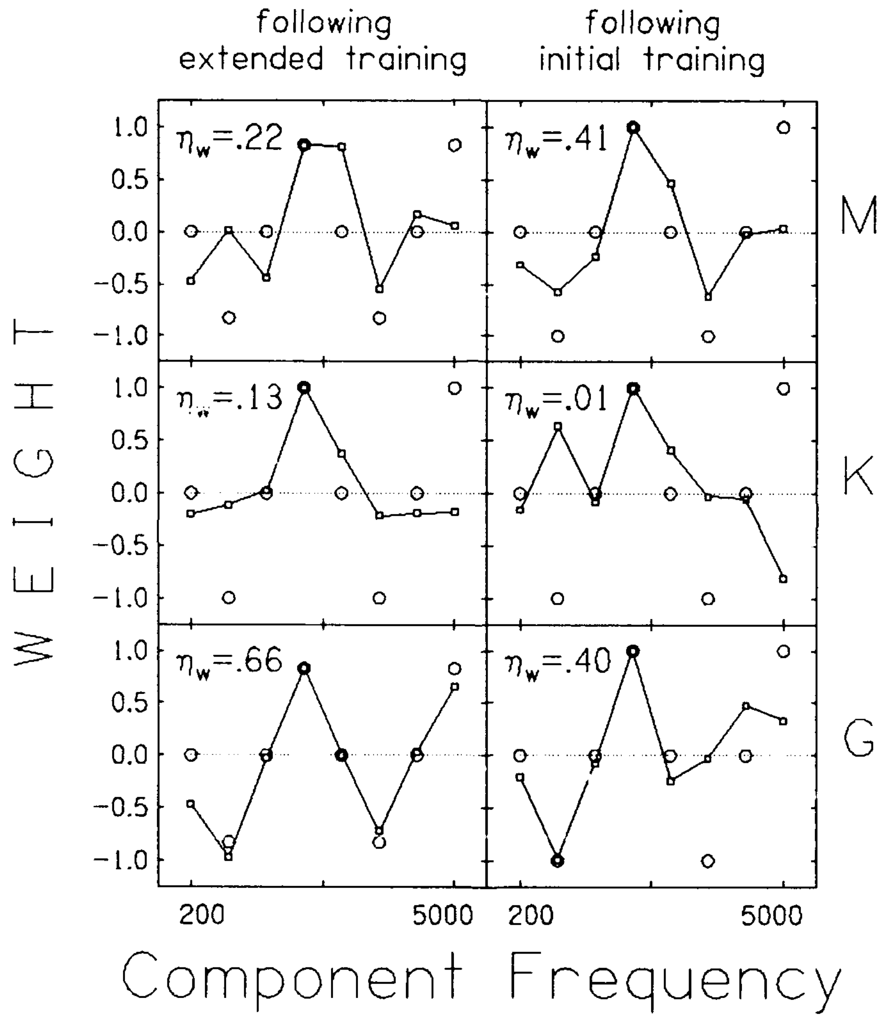
Berg, B.G., and Green, D.M. (1990). Spectral weights in profile listening. J. Acoust Soc Am., 88, 758-766.

Berg, B.G., and D.M. Green (1991). Discrimination of complex spectra: spectral weights and performance efficiency. In Y. Cazals, L. Demany, and K. Horner (Eds.), Auditory Physiology and Perception, 9th International Symposium on Hearing. Pergamon Press (in press).

Durlach, N.I., and Braida, L.D., and Ito, Y. (1986). Towards a model for discrimination of broadband signals. J. Acoust. Soc. Am., 80, 63-72.

Feth, L.L. (1974). Frequency discrimination of complex periodic tones. Percept. Psychophys., 15, 375-378.

FIG. 1


following
extended training
following
initial training
η_w=.22
η_w=.41
η_w=.13
η_w=.01
η_w=.66
η_w=.40
1.0
0.5
0.0
-0.5
-1.0
M
K
G
WEIGHT
200
5000
200
5000
Component Frequency